[redacted] 403

USPS-T-2

BEFORE THE
POSTAL RATE COMMISSION
WASHINGTON, D.C. 20268-0001

EXPERIMENTAL FEES FOR NONLETTER-SIZE BUSINESS REPLY MAIL, 1996

Docket No. MC97-1

DIRECT TESTIMONY OF

LESLIE SCHENK

ON BEHALF OF

UNITED STATES POSTAL SERVICE

TABLE OF CONTENTS

Direct Testimony

of

Leslie M. Schenk

Autobiographical Sketch

My name is Leslie M. Schenk. I am a Senior Economist with Christensen Associates, which is an economic research and consulting firm located in Madison, Wisconsin. I have been employed at Christensen Associates since June, 1995. During my tenure at Christensen Associates, I have worked on many research projects for the U.S. Postal Service.

In 1982 I received a B. A. from SUNY College at Buffalo, with a major in economics and a minor in mathematics. I received an M.A. in economics, and an M.A. in mathematics (with a concentration in statistics) from Indiana University in 1984 and 1986, respectively. In 1995 I received a Ph.D. in economics from Michigan State University.

From 1985 to 1986 I was a research assistant on the economic forecasting modeling project at the Indiana University Business School. There I was responsible for quarterly economic forecasts for industry clients. From 1986 to 1989 I was a demand analyst for Indiana Bell Telephone Company. Among my duties there, I helped prepare analyses for rate case filings before the Public Service Commission of Indiana. I also provided in-house statistical consultation. From 1993 to 1995 I worked as a research assistant at the Institute for Public Policy and

Social Research at Michigan State University. My research there was on nonprofit organizations. From 1983 to 1993, I taught numerous economics, business statistics, and mathematics courses. My recent research for the Postal Service has involved a number of in-field surveys to support Dockets No. MC95-1 and MC96-2.

I. Purpose of Testimony

The Postal Service accounts for Business Reply Mail (BRM) using several different methods. The purpose of my testimony is to present an analysis of two of these methods, weight averaging and reverse manifest. At the request of the United States Postal Service, Christensen Associates has studied three through-the-mail film developers that each use one of the BRM methods under analysis. We were asked to estimate the cost of daily rating and billing procedures associated with these two methods for the three mailers studied. In addition, we were asked to estimate the cost of the sampling procedures associated with determining the postage per pound rate used in weight averaging. We were also asked to evaluate the statistical validity of the sampling procedures used in each of these methods, recommend changes (if any) needed in the procedures, and estimate the costs associated with weight averaging and reverse manifest methods with the recommended sampling procedure changes. Since we were not able to observe any procedures involved in the initial set-up of these systems, we have not estimated any set-up costs associated with these methods.

All cost estimates and recommended sample procedures reported here are specific to these three mailers, and reflect the experience each Postal Service site has in the BRM-related procedures investigated. These cost estimates also reflect the current level of proficiency Nashua has in their manifest system. Any changes in the current procedures or manifest proficiency levels will affect the cost estimates for these methods. From the limited information available on these two

methods of rating and billing BRM, we are not able to determine the applicability of these cost estimates to other firms' BRM processed using these methods.

II. Introduction

This testimony presents cost estimates for three through-the-mail film processors that use Business Reply Mail to receive exposed film from customers. The firms represented here are Mystic Color Lab ("Mystic"), Seattle FilmWorks, Inc. ("Seattle FilmWorks"), and Nashua Photo Inc. ("Nashua"). A weight averaging method is used to rate and bill nonletter-size BRM for Mystic and Seattle FilmWorks. BRM for Nashua is rated and billed using the reverse manifest method. The procedures used for each method are described in detail in witness DeMay's testimony (USPS-T-1).

The analysis presented here is based on site visits made by me and other Christensen Associates personnel, during which Postal Service operations related to these three mailers were observed. We collected sample data during the site visits, and Postal Service personnel also recorded data for a two-week study that we conducted of weight averaging and reverse manifest Postal Service operations.

My testimony covers several major topics. In Section III, I describe our data collection efforts, costing methodology, and the estimated costs of current operations. I discuss the (statistical) reliability of the sampling procedures used in the weight averaging and reverse manifest methods in Section IV, and present our

recommendations to improve the sampling procedures. The impact of the proposed changes in sampling procedures on cost is discussed in Section V.

III. Costing Methodology

The methodology used to derive the costs of rating and billing nonletter-size BRM for Mystic, Seattle FilmWorks, and Nashua is presented in this section. First, data collection efforts are described. In the second part of this section, the general costing methodology and assumptions are discussed. Estimated costs for the current procedures associated with weight averaging and reverse manifesting are discussed in the last part of this section.

A. Data Collection

Data collection was done in two stages. First, data were collected during site visits to Postal Service facilities in Seattle, WA, New London, CT (for Mystic Color Labs), and Parkersburg, WV (for Nashua). Postal Service operations were also observed at Nashua's plant in Parkersburg. In addition, a cost study was conducted at each site.

Site visits were made to each location where the three firms' BRM is rated and billed in order to obtain detailed information on the procedures used for the weight averaging and reverse manifest methods. At each site, Postal Service personnel were interviewed on all phases of BRM-related procedures for each firm. Daily operations were observed. Data were also collected by myself and other Christensen Associates personnel on randomly selected BRM pieces. The visits to

New London and Seattle were timed to coincide with the periodic sampling (i.e., the sampling done to determine the postage per pound conversion rates used in the weight averaging procedure). These sampling procedures were observed and timed.

Based on information obtained during our site visits, and additional information we obtained from Postal Service personnel, a survey of BRM-related activities for the weight averaging and reverse manifest methods was developed. This survey collected data for a two-week period on the time it took each Postal Service employee involved in the weight averaging and reverse manifest methods to complete each associated task. These times were self-reported by each employee, and were reviewed for completeness and accuracy by the employee's supervisor. Once received by Christensen Associates, the surveys were reviewed for any discrepancies or anomalies, which were resolved through discussions with the supervisor. The survey was conducted from October 9, 1996 through October 22, 1996.[1]

The survey forms and instructions are presented in USPS Library Reference EBR - 1. Instructions for survey completion were discussed with supervisors at each site during the site visit, as well as through telephone contact the day before the survey period began, to ensure that only times associated with the rating and billing of nonletter-size BRM for the firms under study were included in the reported times.

[1] Personnel at Mystic were not able to start the survey until October 10th, but continued the survey on October 23rd so that two complete weeks of data could be collected.

B. Methodology and Assumptions

Cost estimates for Mystic and Seattle FilmWorks for daily activities associated with the bulk weighing of daily volumes and with BRM accounting procedures were based on the data obtained from a two-week cost study. This cost study was also the source of data for cost estimates for Nashua for daily sampling, manifest verification, and accounting. Cost estimates for the periodic sampling procedures (i.e., those used at the weight averaging sites to determine postage per pound conversion rates) were based on data collected by myself and other Christensen Associates personnel during the site visits, supplemented by information we obtained from Postal Service personnel.

The hourly wage rates used to cost out the procedures are for FY 1995. The wage rates used in this analysis represent the average wages and benefits for all personnel in the appropriate segment, and are derived using standard CRA methodology. FY 1995 wage rates are used because they are the latest available; no test year time frame has been established for this case.

Standard overhead and piggyback factors used in the CRA are applied to labor costs. The factors used for this analysis are the mail processing overhead factor and the Bulk Mail Acceptance Unit piggyback factor. The sampling and manifest verification procedures for Nashua are performed at the firm's plant, and therefore are treated in this cost analysis like other work done at detached mail units. The appropriate piggyback factor to use in this case is that for Bulk Mail Acceptance Unit (as there is currently no separate piggyback factor calculated for detached mail

units or plant-loads). It is anticipated that most reverse manifest systems will require sampling and manifest verification at the firms' plants, since this mail stream will probably bypass local USPS plants, and since verification will require access to the participating firm's manifest system in electronic form. For consistency across all nonletter-size BRM costs, and since the same mailflows are involved, the BMAU piggyback factor is also applied to costs for Mystic and Seattle FilmWorks.

Volumes used in cost estimates for Mystic come from estimates obtained during the site visit, since no records are kept by the Postal Service on the daily number of nonletter-size BRM pieces received for Mystic. Average daily volume was estimated by inflating volumes from sample sacks to the total number of sacks received for each sample day. Annual volume was estimated by inflating average daily volume by a factor of 300 (to represent the average number of days per year that mail is processed for Mystic).

Volumes used in cost estimates for Seattle FilmWorks were obtained from PERMIT records for the three Seattle FilmWorks accounts[2] for the time period that the cost survey was conducted. These volumes are estimates derived by the Postal Service using the piece weight distributions obtained during periodic sampling. The Postal Service facility which rates and bills Seattle FilmWorks' BRM is the only site

[2] Seattle FilmWorks receives nonletter-size BRM under three separate "accounts." Separate permit and accounting fees are paid for each. Mailpieces for each account are not commingled in the incoming mailstream arriving at the Seattle facility where Seattle FilmWorks' BRM is rated and billed, nor are nonletter-size and automatable pieces for the same account commingled. We were told that Seattle FilmWorks maintains these separate accounts for marketing or internal accounting purposes.

of the three in this study which enters daily activities for the film processor in the PERMIT system.

Volumes used in cost estimates for Nashua were obtained from volumes reported in the cost study. These daily volumes are reported by the firm to the Postal Service and recorded on PS Form 8159 (See Docket No. MC96-3, USPS LR-SSR-148, pp. 105-108). Annual volumes were estimated by multiplying the average daily volume by the number of days that mail is processed at Nashua (360 days).

C. Estimated Costs of Current Operations

Using the data collected during the site visits on sampling times, we are able to estimate a manual (baseline) cost per piece for nonletter-size BRM. This is an estimate of the cost to manually weigh and rate each BRM piece (i.e., the procedure used prior to the adoption of weight averaging or reverse manifesting were used for these mailers). The estimated cost[3] per piece of manually weighing and rating each piece (including accounting tasks) is given in Exhibit USPS-T-2A.

The time per piece to manually count and rate was calculated by taking the total time per day to sample and rate pieces as we observed during the periodic sampling for Mystic and Seattle FilmWorks, and dividing by the number of sample pieces. The estimated time per piece was multiplied by average hourly rates for the segment which performs the daily tasks associated with the firms' mail, and also multiplied by standard overhead and piggyback factors. As shown in Exhibit USPS-

[3] All detailed cost derivations are given in my confidential workpaper.

T-2A, the average cost per piece[4] of manually rating and billing nonletter-size BRM is $0.1147 for these mailers.[5]

The daily sampling procedures done by the Postal Service that we observed at Nashua differ substantially from those which would be used under a manual counting and rating system. In the manifest system, detailed information on each piece (e.g., customer number, ZIP Code) is recorded by Postal Service personnel so that the sample piece can be identified in the company's manifest. Under a manual system, this detailed information would not have to be recorded. Therefore, inclusion of the current sampling time at Nashua in this calculation would overestimate the costs of manually rating this mailstream.

It is reasonable to assume that a manual rating system would not vary substantially across sites, so the average cost calculated above is representative of any site using manual procedures to rate each piece of nonletter-size BRM.

The estimated costs of current BRM procedures for the three film developers are given in Exhibit USPS-T-2B. Daily costs were determined by taking the average daily time (in hours) needed to complete each activity, and multiplying by the appropriate hourly rate. We then multiplied this result by standard overhead and piggyback factors. The cost per piece for daily weighing was obtained by then dividing by the estimated average daily mail volume. The daily weighing costs per

[4] The per piece cost is reported because costs of manually rating this mail are a function of the volume of this mail received.

[5] It should be noted that this is not the current cost of rating and billing nonletter-size BRM for these mailers, as they are not currently manually rating and billing this mail. This figure is provided for comparison purposes only, not to indicate costs associated with current practices used for these mailers.

piece[6] are $0.0151 for Mystic, and $0.0104 for Seattle FilmWorks. The monthly accounting costs are derived by multiplying the daily accounting costs by 25 (the average number of days per month that mail is weighed and rated for Mystic and Seattle FilmWorks). Monthly accounting costs for Mystic and Seattle FilmWorks are $893 and $162 per account,[7] respectively. Accounting costs differ between the two weight averaging sites because of different accounting procedures used – a manual procedure is used for the Mystic accounting, while accounting for Seattle FilmWorks is done through the (automated) PERMIT system.

For Mystic and Seattle FilmWorks, costs associated with periodic sampling were derived by taking the time (in hours) needed to collect sample data and calculate new conversion factors, and multiplying by the appropriate hourly rate. Estimated sample training costs[8] are included for Seattle FilmWorks. At the Postal Service facility where Seattle FilmWorks' BRM is rated and billed, training is done each time a sample is taken because sampling is done infrequently, and different clerks take the sample each day of the sample week. Personnel are rotated because of unique conditions[9] in the work area. Training is done on the workroom floor, and mailpieces are handled as part of the training, so these costs are attributable.

[6] The per piece cost is reported because costs of the daily bulk weighing are a function of the volume of mail received.
[7] Monthly costs are reported because accounting costs are not a function of the volume received.
[8] Estimated training times were obtained from the supervisor responsible for the training.
[9] Daily bulk weighing and periodic sampling of Seattle FilmWorks BRM are done in an area just inside the dock platform. In this area there is a very noisy air compressor. In addition, it was reported to us that rats, feral cats, and even raccoons have been spotted in the warehouse area which is colocated with the area where the bulk weighing and sampling of this mail is done.

We then multiplied by standard overhead and piggyback factors. As Exhibit USPS-T-2B shows, the cost of sampling for Mystic is $1,151 per sample, and for Seattle FilmWorks is $241 per sample per account.[10] The cost of sampling is higher for Mystic than for any of the Seattle FilmWorks accounts for two reasons: more sample pieces are currently drawn for Mystic than for any of the Seattle FilmWorks accounts, and there is a difference in the sampling and calculation procedures used for Mystic and Seattle FilmWorks, as described in Part A of Section IV.

In the past the sample for Mystic has been drawn and the conversion factors calculated by the New London postmaster. We have been informed that a supervisor is being trained to perform these tasks in the future. The total cost of sampling for Mystic Color Labs, assuming that a supervisor does the sampling and calculations, is $1,265 per sample, as shown in Exhibit USPS-T-2B.

The total monthly estimated cost for the reverse manifest system at Nashua, given current sampling and accounting procedures, is $4,053, as shown in Exhibit USPS-T-2B. This cost estimate was derived in a similar manner to those for the weight averaging sites. Since daily sampling, manifest verification, and accounting procedures are not a function of the volume received, a per piece cost is not reported for reverse manifesting.

[10] Sampling costs are a function of the sample size (which is fixed, at least over the course of the experiment), not a function of volume received.

IV. Reliability of Sampling

In order to determine whether the correct sampling procedures are being used to obtain an accurate postage per pound conversion factor for the weight averaging method for Mystic and Seattle FilmWorks, the sampling methods were observed, and samples of pieces were taken by myself and other Christensen Associates' personnel. In addition, daily sampling procedures used at Nashua were observed, and sample data collected by Postal Service personnel were obtained. We then used standard statistical methods to analyze the sample data and procedures used. The sampling procedures used at each site are described first, with analysis and recommendations following.

A. Analysis of Sampling Procedures at Weight Averaging Sites

1. Current Sampling Procedures

The sampling procedures used for Mystic and Seattle FilmWorks are similar, in that a version of two-stage sampling is used and the sample period is one week long (5 consecutive days). For Mystic, two sacks are selected each day of the sample period (10 sacks in total), and all the pieces in the selected sacks are sampled. For Seattle FilmWorks, they sample 200 pieces per account per day of the sample (1,000 pieces in all) by selecting a (relatively full) sack at random for each account, and selecting the first 200 pieces from that sack. For one of the Seattle FilmWorks' accounts, the sacks tend to be less full, so on some sample days more than one sack for that account has to be sampled to get the target 200 sample pieces.

The sampling procedures differ for Mystic and Seattle FilmWorks in terms of what data are recorded. Individual pieces weights are recorded for Mystic, whereas counts by ounce increment (e.g., the number of pieces over one ounce in weight but less than two ounces) are recorded on sample pieces for each Seattle FilmWorks account. For Mystic, a postage per pound rate is calculated from the sample data. This rate is multiplied by the daily total weight of all pieces to obtain the total postage due. For Seattle FilmWorks, a pieces per pound rate is calculated for each account from the sample data. This rate is multiplied by the total daily weight for that account, to arrive at an estimated total daily piece count. The distribution of pieces per ounce increment per account is also calculated from the sample data. This distribution is applied to the estimated total daily piece count per account, to get an estimate of pieces in each ounce increment. These piece counts are entered into the BRM module of the PERMIT system, and the postage due per ounce increment and total postage due are automatically calculated, and a bill produced. The postage due calculations used for Seattle FilmWorks are different than that used for Mystic because the Postal Service facility which rates and bills Seattle FilmWorks uses the BRM module in the PERMIT system to record all Postage Due activities.

2. Analysis of Sample Procedures and Recommendations

For both Mystic and Seattle FilmWorks we drew sample pieces using a systematic sampling method. Every fifth piece was selected from every nth sack, where the sack sampling rate for each site was determined based on expected

volume and time constraints. We recorded each sample piece's weight, as well as corresponding sack information (including the total weight of the sack, and the dispatch from which the sack was sampled). The total number of pieces sampled was 1,915 and 1,309 for Mystic and Seattle FilmWorks, respectively.

The mean and variance of postage per pound estimates by sacks for Mystic and Seattle FilmWorks are shown in Exhibit USPS-T-2C. These results show that the estimated postage per pound rate varies considerably across sacks. This variation across sacks indicates that the current sampling method, where all pieces from a few sacks are sampled (rather than sampling some pieces out of many sacks), may be inefficient (in the statistical sense, i.e., not minimum variance) compared to estimates obtained by sampling random pieces from many sacks.[11]

In order to estimate the correct sample size needed to obtain a given precision level for total postage due, the variability of postage per pound is required. A bootstrap procedure was used to measure this variability. A detailed description of the bootstrap methodology is given in Appendix A.

a) Mystic

The bootstrap results for Mystic are given in Exhibit USPS-T-2D. The first line in Exhibit USPS-T-2D shows the results if the sample is drawn randomly from all sacks. For example, if there is no seasonality in postage per pound (i.e., the distribution of the weight per piece does not fluctuate throughout the year in a

[11] If the estimates are inefficient, then the variance of the estimate is not as small as it could be, which means that the estimate of adjustment factors is less precise than (not as accurate as) it could be, at the 95 percent confidence level. This has ramifications for revenue leakage.

regular pattern), then 1,000 sample pieces per month will obtain an estimate of the monthly revenue[12] to within 2.3 percent of the true value with 95 percent confidence. If a sample of 1,000 pieces were taken once a month, for a total of 12,000 sample pieces per year, then the estimate of total annual postage due would be within one percent (0.66%) of the true value with 95 percent confidence.

If 5,000 pieces were randomly selected per year from all BRM received, the resulting estimate of annual postage due would be within 1.02 percent of the true value, with 95 percent confidence. Currently, 5,000 sample pieces are drawn for Mystic per year, but not randomly from all pieces. The sample is drawn for Mystic only once a year, and a two-stage sampling procedure is used, as described above. Two sacks are selected at random each day for five days (a total of 10 sacks altogether), and all pieces are sampled from these selected sacks. In the second row of Exhibit USPS-T-2D, it is shown that the precision level is 2.86 percent for the postage due estimate obtained from a complete enumeration of ten sacks.

In general, the reliability of the estimates from a two-stage sample procedure is lower than for a random selection from all pieces, for the same sample size. In addition, precision increases with the number of sacks sampled, holding the number of pieces sampled constant. An example from the samples taken for Mystic illustrates these two points. For Mystic, there were on average 350 pieces per sack in our sample. Therefore, a two-stage sample with 13 sacks drawn, and all

[12] Daily postage due is obtained by multiplying the postage per pound by the weight of all sacks of BRM received each day. Since the weight of all sacks is non-stochastic (i.e., non-random or known with certainty since all sacks are weighed), the precision level for the estimated postage due is the same as the precision level for the estimate of postage per pound.

pieces in those sacks sampled, yields approximately 5,000 pieces per sample. The precision level for this two-stage sample is 2.49 percent, whereas for a sample of 5,000 pieces drawn randomly from all pieces the precision level is 1.02 percent. If 65 sacks were sampled, with a 1 in 5 piece skip, the sample size would be 5,000 pieces and the precision level would be 1.46 percent. As these results indicate, for a given sample size random sampling from all pieces provides the most accurate estimates. However, a two-stage sampling procedure is more practical to implement and takes less time (holds up processing less) than a random selection from all pieces. The precision level for two stage sampling approaches that of random sampling, as more sacks are sampled. A five-day sample procedure where more than two sacks per day are sampled from will improve the reliability of the estimated adjustment factors, thus providing more assurance that correct postage is collected. The sample design recommended for the experiment for Mystic is discussed in Part (d.) of this Section.

b) Seattle FilmWorks

Exhibit USPS-T-2E shows the bootstrap results for Seattle FilmWorks. These results can be interpreted in the same way as described for the Mystic results. Bootstrapping for Seattle FilmWorks was done on samples from each account separately.

Currently for Seattle FilmWorks, two hundred pieces from each account are sampled for five days (1,000 pieces in total from each account). This sampling is currently done once a year. As shown in Exhibit USPS-T-2E, random draws of

1,000 pieces will produce estimates of total monthly postage due which will be within two percent of the true value with 95 percent confidence for each account.

However, this overstates the accuracy of the postage due estimates obtained using the sampling procedures currently in place for Seattle FilmWorks. The 1,000 pieces sampled per Seattle FilmWorks account are not currently drawn randomly from all pieces received. A two-stage sampling procedure is used, as described above. The precision levels for a two-stage sampling, where 5 sacks (one sack per day for five days) are selected and completely enumerated, would yield postage due estimates within approximately 4 percent of the true value for each account, as shown in Exhibit USPS-T-2E. This overstates the accuracy of the actual sample selected, however, since each selected sack is not completely enumerated under the current sampling method used at Seattle – only the first two hundred pieces for each account are selected. Sacks from our sample containing mail for two of the Seattle FilmWorks accounts (accounts 25 and 56) on average contained over 260 pieces per sack, so sampling for these accounts is done by selecting the first 200 pieces from one sack selected at random. For the other Seattle FilmWorks account, the average sample sack contained only 184 pieces, so the first 200 pieces from two selected sacks are generally sampled. It should be noted that Postal Service personnel stated that heavier pieces could fall to the bottom of the sacks during handling and transporting; we are not able to test this theory with data currently available.

c) The Effects of Seasonality on Precision Levels

The samples used to produce the bootstrap results reported in the preceding sections could be used to produce inferences on the precision levels of annual postage due estimates if the samples were selected at a time representative of the whole year, and there was no seasonal pattern in weight per piece. Our samples for Mystic and Seattle FilmWorks were taken in late September-early October, which has been described to us by firm and Postal Service personnel as a time of year when there are fewer heavier pieces received. If this is correct, the average weight per piece in our sample is lower than would be the case at other times of the year.

Seasonality would have two effects on our analysis of sampling reliability. First, the estimate of postage per pound reported above would be a biased estimate of the average annual postage per pound estimate. Also, the variance in postage per pound would be higher than is reported here. If there is seasonality in the weight per piece, and fewer heavier pieces are received at the time our samples were drawn, then the variances reported in Exhibits USPS-T-2D and 2E represent lower bounds on the variance at any given time in the year, and therefore a lower bound on the precision level of the estimates (i.e., the estimates are actually less precise than reported here). If there is seasonality, then the mean postage per pound reported in Exhibits USPS-T-2D and 2E is a biased estimate of the mean annual postage per pound.

Firm and Postal Service personnel at Mystic, and Postal Service personnel at Seattle, reported that more heavy pieces[13] are received during the summer months and immediately after major holidays. Seasonality will cause the postage per pound conversion factor to differ throughout the year, and also cause the variance to be higher than reported here. If there is seasonality, then the sampling procedures currently used (i.e., drawing one sample per year) produce biased estimates of the average postage per pound, with less precision. Less precision increases the possibility of revenue leakage.

Since no data are currently available to us to determine the seasonal pattern (if any) for these mailers, we were not able to quantify the effect seasonality has on bias or precision. A trial period, where weight per piece data are collected monthly for at least one year, using statistically-valid sampling procedures, would provide the data needed to determine the proper sample design, including the periodic sampling time (monthly, quarterly, or other).

d) Trial Sample Design for Mystic and Seattle FilmWorks

A two-stage sampling method is recommended for the trial: each month for a year select 20 sacks at random per account, and sample all pieces in the selected sacks. With a sample of 20 sacks per month per account, the estimated postage due will be within approximately two percent[14] of the true postage due for each

[13] These heavier pieces include one-use cameras and envelopes containing more than two rolls of film.

[14] For Mystic the precision level is 2.05 percent, as shown in Exhibit USPS-T- 2D. For the Seattle FilmWorks accounts the precision levels are 2.1, 1.7, and 2.0 percent, as shown in Exhibit USPS-T-2E.

month for which the variance in weight per piece is similar to the time sampled. That is, with 95 percent confidence we can say that the resulting estimated postage due for Mystic and for each Seattle FilmWorks account will be within 2 percent of the true value, for any month with mail flow patterns similar to our sample period.

This sample design provides for some revenue protection and allows for more data to be collected to determine the seasonal patterns, without over-burdening Postal Service personnel's workload. In the months during the experiment where the piece weights are more variable than our sample period, the estimated postage due will be less precise (than the two percent reported above), but should still be in a reasonable range. The suggested sampling procedure also controls for possible variation across sacks (e.g., because of possible differences in geographic distribution), as well as variation within sack (e.g., heavy pieces falling to bottom of sack). More importantly, with a monthly random sample the seasonal pattern can be determined, and the sample design adjusted accordingly based on this new information.

In order to spread out the workload and account for daily fluctuations, we recommend that 4 sacks per day be sampled per account for a week[15] (five consecutive days) each month, where the week selected varies by month.

If weight per piece data were collected for Mystic and Seattle FilmWorks for one year, the annual reliability of total postage due could be re-assessed for the

[15] Since one of the Seattle FilmWorks accounts does not typically receive more than 2 sacks per day, sampling for that account would have to be done over a two-week period each month.

observed seasonal pattern. The sampling procedure can then be redesigned to reflect these new data collected on seasonal patterns.

This discussion on the reliability of sampling with a weight averaging system has focused on the systems in place for Mystic and Seattle FilmWorks. For any other implementation of weight averaging, the sample design would depend on factors specific to each firm relating to mail flows (e.g., variation in piece weights).

B. Analysis of Sampling Reliability for Nashua

1. Current Sample Procedures

The daily sampling procedures at Nashua consist of selecting 50 sample pieces (30 in the morning, 20 in the afternoon) and recording the piece weight as well as identifying information: unique customer number (or street address if no customer number available), ZIP Code, and Nashua's "media code". Each piece is then rated. It is this actual piece postage that is compared with the manifest postage, when the piece is subsequently located in the firm's manifest.

The fifty sample pieces are selected randomly from all pieces that are in a bin in a room near the dock entrance. In this room the incoming sacks are weighed, and the pieces are sorted by Nashua employees. Nashua employees informed us that their culling procedure includes separating out envelopes with only reprints (very light, one-ounce standard pieces) and envelopes with one-use cameras (heavy pieces, generally 5 ounces or over). These pieces are culled out of the main in-plant mailstream to facilitate processing within the plant, according to the personnel from Nashua with whom we talked. The Postal Service samples are selected from

this bin after heavy and very light pieces are culled out.[16] The Postal Service sample is therefore taken from a censored population of all BRM pieces Nashua receives, and so is not a random sample (i.e., all BRM pieces do not have an equal probability of being selected). Since the sample is not random, estimates from this sample could be biased (i.e., not representative of the adjustment factor for all Nashua pieces). Since total postage (from all pieces received) is adjusted on a daily basis using the adjustment factor determined from the sample pieces, it is important that the sample be randomly drawn from the population of all pieces.

2. Analysis and Recommendations

We drew sample BRM pieces from sacks at the Parkersburg Post Office, and from sacks which were directly transported to the Nashua plant from other Post Offices. We used a systematic sampling method, whereby we selected every nth sack, and every 10th piece within each selected sack. The sampling rates used were determined based on expected volumes and time constraints. For each piece selected, we recorded the weight of the piece, as well as other information needed to look up the piece in the company manifest (e.g., customer identification number, ZIP code), i.e., we recorded the same information that is recorded by the Postal Service clerks as part of their daily routine.

Our sampling procedure, however, differs from that used by the Postal Service. We sampled pieces directly from the sacks of mail the firm receives, before any culling or processing by the firm. The population we sampled from

[16] Because the Postal Service clerk arrives after the culling process has taken place.

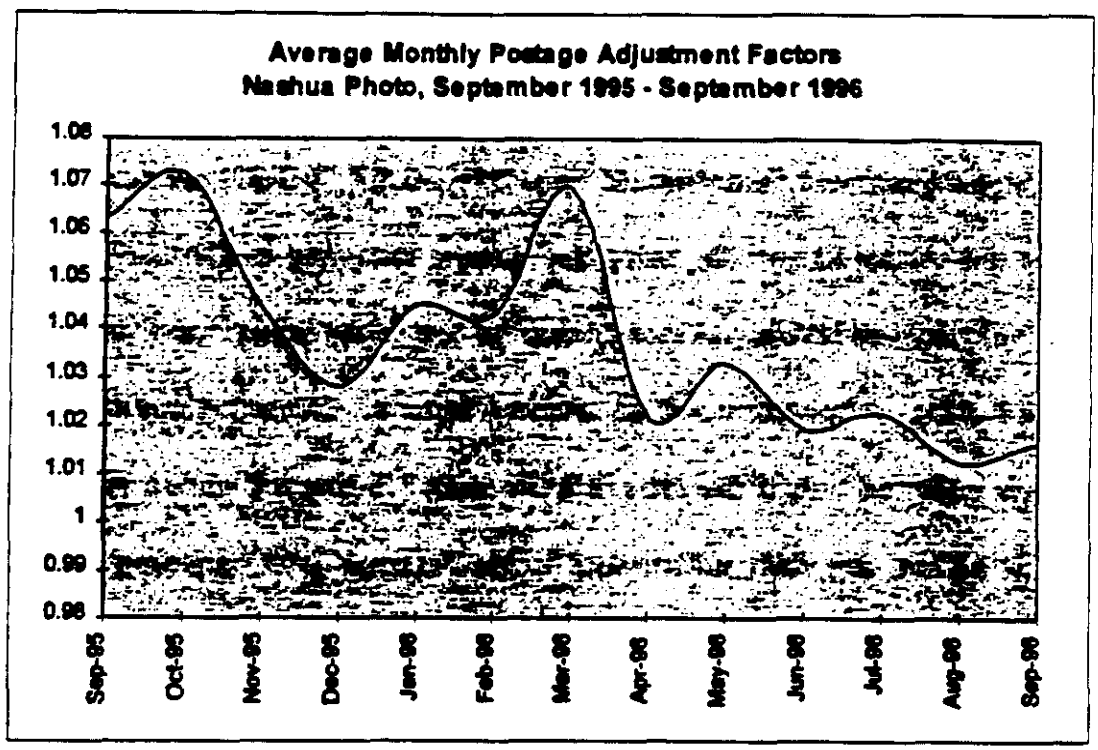
includes all pieces received by Nashua on the day we sampled (September 24, 1996).

The Postal Service clerk informed us that they always sample from the bin *after* the culling process. Our sample pieces were drawn from the uncensored population, whereas the Postal Service sample is drawn from the censored population.

The Postal Service clerk verified the postage on the 110 pieces we drew from the uncensored population. In addition, we obtained data on a limited number of pieces from the uncensored population selected and verified by a Postal Service clerk (not as part of the regular sampling process). The number of pieces from the uncensored population for which we could obtain manifest revenues is limited, so no statistical inference can be drawn on the accuracy of the manifest for the uncensored population. The actual and manifest postage did not agree for 25 percent of the 110 sample pieces that were verified from the uncensored population. Fifteen pieces (14 percent) of the 110 sample pieces were heavier pieces. The actual and manifest postage agreed for only one-third of the heavier pieces.

We also obtained data on the Postal Service's samples for one year (September 1995 - September 1996). The average monthly postage adjustment factor over this time period showed a downward trend. Witness DeMay informed us that this downward trend could be attributed to a learning curve (i.e., the firm improving piece postage calculation procedures). Looking at the average

adjustment factors by month for September 1995 - September 1996, as shown in the chart below, we can see that the trend in adjustment factors levels out by July 1996. For this reason, all bootstrapping analysis was performed on daily postage due (actual and manifest) data for July - September, 1996.



A bootstrap procedure[17] was used to determine the confidence interval and precision level for the annual postage adjustment factor. This bootstrap procedure produced a distribution of bootstrapped annual adjustment factors. The 95 percent confidence interval was calculated directly from this distribution. The results of this bootstrap procedure for Nashua are presented in Exhibit USPS-T-2F.

[17] The bootstrap procedure is discussed in detail in Appendix A.

These bootstrap results show that, given the current sample size (50 per day), the average postage adjustment factor is within (-0.25, +0.28) percent[18] of the true value (for the censored population) with 95 percent confidence, on an annual basis. That is, over all the daily adjustment factors for a year, the average adjustment factor would be within (-0.25, +0.28) percent of the true value for the censored population, based on daily samples of 50 pieces selected from the censored population. For all daily adjustment factors over a representative month, the average adjustment factor would be within (-0.87, +0.97) percent of the true value for the censored sample, using the current sampling methodology.

The precision levels reported above for the mean adjustment factor were calculated using all bootstrapped adjustment factors. Currently, the manifest postage is adjusted only if the ratio is more than 1.5 percent above or below 1.000. In effect, any daily adjustment factor between (and including) 0.985 and 1.015 is currently treated as if it were equal to 1.000. We also bootstrapped the ratio adjustment factors to follow this procedure (i.e., where all daily adjustment factors between 0.985 and 1.015 were changed to 1.000 before calculating the mean and precision level). The results of this second bootstrap procedure show that the practice of adjusting postage on a subset of days produces biased annual average adjustment factor estimates. As shown in Exhibit USPS-T-2F, the postage adjustment factor is underestimated. The average adjustment factor is 1.0186 if

[18] The distribution of adjustment factors is asymmetric (there are more adjustments up than down), and there is more variation in the adjustment factors greater than 1.000. This accounts for the asymmetry in the precision levels.

postage is adjusted only when the adjustment factor is more than 1.5 percent above or below 1.000 (current practice), but is 1.0197 if postage is adjusted every day regardless of the level of the adjustment factor. The bias in the average adjustment factor is due to the fact that more daily adjustment factors fall between 1.000 and 1.015 than fall between 0.985 and 1.000.[19] On average, we estimate that postage due from Nashua when adjustments are not made daily is underpaid by over $5,000 per year. This bias can be reduced with improved data entry methods at the firm.

To determine postage due for Nashua's BRM, the daily adjustment factors are currently applied to total manifest revenue, even though the adjustment factors are determined based on the censored sample. This would be an acceptable procedure if one could assume that the censored sample the Postal Service draws has the same ratio adjustment factor as would a sample from the whole population. As discussed above, limited information is currently available on the difference between per piece actual and manifest postage for very light and very heavy pieces (i.e., pieces less than one ounce or greater than 4 ounces). Based on the pieces we sampled, and the heavy and light pieces sampled by the Postal Service, there is some evidence that the inclusion of heavy pieces would change the adjustment factor.

[19] If the adjustment factors were evenly distributed in the "no adjustment" interval (0.985 to 1.015), then their inclusion would not affect the postage due to the Postal Service, as their mean would be 1.000.

Since *total* postage due is adjusted, sample pieces should not be drawn from the censored sample exclusively, unless it is statistically determined that the postage adjustment factor is the same for the uncensored sample as it is for censored sample.

To develop the correct sample design to obtain precision in estimates of the total BRM postage due for Nashua, we need information on the variance of total adjusted postage due, calculated using adjustment factors derived from uncensored samples. Since these data are not currently available to us, we recommend that at least 10 pieces per day from the light and heavy pieces be sampled for one year so that enough data can be collected that the accuracy of the manifest for this part of distribution can be determined. The cost of taking this larger sample is incorporated into the cost estimates presented in Section VII. Since no definitive data are available to assess the manifest performance for these light and heavy pieces, we cannot estimate the impact on Postal Service revenues of including these pieces in the postage adjustment factor calculation. Data could be available from the experiment to make this assessment.

One additional procedural change recommended is to change the way sample pieces selected by the Postal Service are re-entered into Nashua's processing stream. Currently, the Postal Service clerk will take the tub of sample pieces and either give it directly to one of the Nashua clerks, or place it on a staging shelf for the next available clerk. The instructions given the postal clerks however were to put the sample pieces back into the culling bin with the rest of the BRM pieces to

be processed. We recommend that the clerks follow the original instructions. This practice has less potential for bias and correlation within a day's sample observations. The manifest system should be tested daily across all Nashua data entry personnel, not just selected ones.

This discussion on the reliability of sampling with a reverse manifest system has focused on Nashua, which is the only example of a reverse manifest system used for BRM. For any future reverse manifest system, the sample design would depend on factors specific to each firm, such as mail flows (e.g., variance in piece weights) and manifest accuracy.

V. Estimated Costs with the Recommended Sampling Procedures

Exhibit USPS-T-2G shows the cost estimates[20] with the recommended sampling procedures for Mystic and Seattle FilmWorks, respectively. It is recommended for the first year that 20 sacks per month are sampled, with the weights of all pieces in the selected sacks recorded. As shown in Exhibit USPS-T-2G, the monthly fixed cost for Mystic, which includes the monthly accounting and sampling costs, would be $3,424, while the variable costs (the per piece cost of daily weighing) is $0.0151. The monthly fixed costs for Seattle FilmWorks is $902 per account, while the variable costs are $0.0104 per piece. These cost estimates were derived assuming that the calculations on the sample data are done manually, as they are currently done.

[20] Detailed derivations can be found in my confidential workpaper.

Exhibit USPS-T-2G also shows the cost per piece for Nashua, with the recommended extended sample procedures (sample heavy and light pieces), as well as the accounting changes recommended. The total cost per day for sampling, manifest verification, and accounting with the recommended procedures is $164 per day ($4,908 for a 30-day month).

The sample information for both Mystic and Seattle FilmWorks are currently recorded and analyzed manually. The cost of doing this analysis manually is greater than if these procedure were computerized. Computerization would also make the test year evaluation of seasonal patterns and sampling procedures more cost effective. The Postal Service could secure or develop software that could be used on PCs with scales attached, so that piece weight information could be recorded and analyzed easily. Use of the computerized sampling system would considerably shorten the time required to do calculations for each sample, and make subsequent analyses more efficient. Exhibit USPS-T-2H shows the estimated costs with the automated sampling and calculation procedures. Computerized procedures would not affect sampling time significantly, but would shorten calculation time considerably. We have estimated that calculation and analysis time, using a computerized recording and calculation procedure, would be approximately 60 minutes per sample for Mystic, and 120 minutes per sample for Seattle FilmWorks.[21]

[21] The calculation and analysis time could be lower in the long run, after sites become more experienced with the new software. More information on the impact of computerization on sampling time will be obtained during the experiment.

Using these estimates, the monthly fixed cost per account with computerized calculation is $2,441 for Mystic, and $808 for Seattle FilmWorks.[22]

[22] If we assume that moving to computerized sample recording and calculations would require a capital outlay of $4,500 ($3,000 for computer, $1,500 for scale attached to the computer), then the cost savings to the Postal Service will exceed the capital outlay after less than one year.

Appendix A – Description of Bootstrap Methodology

Confidence intervals for means and proportions can be calculated analytically for most samples. In some instances however, the construction of a confidence interval may be too complex for standard theory to handle, or the appropriate sample results are not available. Bootstrapping[1] is a nonparametric method which uses extensive computing to construct confidence intervals, estimate statistics, or estimate standard errors of regressions.

When a sample is drawn randomly from a population, it is representative of that population; there is a similarity between the sample and the population. The bootstrap procedure in practice is done by sampling a number of pieces with replacement from the sample (which is a representation of the population). The statistic in question (e.g., mean) is calculated for this sample. This procedure is replicated many times; the resulting distribution of the sample statistics is known as the "bootstrap" sampling distribution. The variance of the estimate is based on the distribution of means from the pooled replications. The bootstrap process is repeated until the estimated variance converges. Confidence intervals and precision estimates are derived from this distribution.

Bootstrapping works well in deriving confidence intervals for non-normal distributions, or to estimate complicated parameters. The bootstrap procedure has

---

[1] More information on the bootstrap method can be found in "Better Bootstrap Confidence Intervals," by B. Efron, Journal of the American Statistical Association, 1987, Vol. 82, No. 397, pp. 171-185; "Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy," by B. Efron and R. Tibshirani, Statistical Science, 1986, Vol. 1, No. 1, pp 54-77; Introduction to the Theory and Practice of Econometrics, Second Edition, G. Judge *et al.*, New York: John Wiley and Sons, 1988, pp. 416-419; Introductory Statistics for Business and

been successfully empirically tested with known distributions (see the references cited in footnote 1 for further discussion of the properties of the bootstrap procedure).

For the analysis on sample reliability reported here, we needed to calculate precision level using the bootstrap method rather than analytically, because of the limitations of available sample data. For the samples drawn at each weight averaging site, one postage per pound estimate is derived. This situation falls in the domain of bootstrapping, since only one realization of the value in question is available, and not a distribution of values (and therefore no variation) on which to base an analytical estimate of sample precision. The sample data available on the postage adjustment factors at Nashua are limited as well, because they are drawn from the censored population. In addition, the sample data from Nashua are not normally distributed and, as stated earlier, one of the benefits to using the bootstrap method is that it is successful in dealing with non-normal distributions.

We did our bootstrapping analysis using Fortran programs on a Data General Aviion mainframe computer. Copies of the programs used are provided in USPS Library Reference EBR-2.

For the Mystic and Seattle precision mean postage per pound and precision level estimates, we bootstrapped on the sample pieces drawn during our site visits. Sample pieces were mapped to their source sack, so that we could replicate both random sampling from all sacks, and a two-stage sample. Bootstrap

---

Economics, Fourth Edition, by T. Wonnacott and R. Wonnacott, New York: John Wiley and Sons, 1990, pp. 277-279.

results are reported for 5,500 bootstrap replications. The result of this procedure is a distribution of postage per pound calculations. Precision levels are based on the mean and variance of this distribution.

A bootstrap procedure was used to determine the confidence interval and precision level for the annual postage adjustment factor for Nashua. The data used in this bootstrap procedure were the daily samples taken by the Postal Service for July - September, 1996. Each observation included the actual and manifest postage for each piece, so that the postage adjustment factor could be calculated from any sample drawn from this data set. "Daily" bootstrap samples were constructed by sampling 50 pieces randomly with replacement in each bootstrap replication. Ninety thousand "daily" replications were done, and the daily postage adjustment factor calculated for each replication. The average adjustment factor was calculated over each group of 300 "daily" observations, to replicate annual adjustment factors. This bootstrap procedure produced a distribution of bootstrapped annual adjustment factors. Precision levels are based on this distribution. Since the distribution is asymmetric, the 95 percent confidence interval and the precision levels were derived empirically. That is, from the distribution of bootstrapped annual adjustment factors, it was determined which adjustment factors cut off 2.5 percent probability in each tail of the distribution. These values give the 95 percent confidence interval for the mean annual adjustment factor. The precision level for the mean annual adjustment factor is then derived by determining the percentage difference between the mean

factor and the values which give the upper and lower bounds of the confidence interval.

Exhibit USPS-T-2A – Costs of Manually Rating and Billing BRM for Mystic and Seattle Filmworks

Site	Sample Time per piece (minutes)	Cost of Manual Counting/Rating per piece
Mystic	0.1404	$0.1251
Seattle	0.1174	$0.1044
Average	0.1289	$0.1147

Exhibit USPS-T-2B – Current Costs of Rating and Billing BRM for the Three Film Developers

Firm	Activity	Cost (per piece)	Monthly Cost
Mystic	Weighing	$0.0151	
	Accounting		$893
	Sampling (done by Postmaster; 5,000 sample pieces, from 10 sacks)		$1,152
	Sampling (done by Supervisor; 5,000 sample pieces, from 10 sacks)		$1,265
Seattle (per account)	Weighing	$0.0104	
	Accounting		$162
	Sampling (1,000 sample pieces from 5 sacks, per account)		$241
Nashua	Sampling (50 pieces per day, sampled from censored population) and manifest verification		$3,615
	Accounting		$438
	Total		$4,053

Exhibit USPS-T-2C – Variance in Revenue per Pound, Sample Sacks

Sack	Mystic Revenue per Pound	Seattle–Account 25 Revenue per Pound	Seattle–Account 56 Revenue per Pound	Seattle–Account 63 Revenue per Pound
1	6.57	7.22	7.14	6.62
2	7.11	7.65	7.94	5.95
3	6.77	7.18	7.62	6.38
4	6.54	7.13	7.85	6.77
5	6.68	7.21	7.33	6.92
6	6.21	6.91	7.22	6.76
7	5.96	7.01	7.20	6.64
8	7.18	6.68	7.35	
9	6.33	7.31	7.60	
10	6.37	6.54		
11	6.73	8.23		
12	6.73			
13	6.44			
14	6.65			
15	6.44			
16	6.89			
17	6.55			
18	6.67			
19	6.54			
20	6.91			
21	6.72			
22	6.58			
23	6.79			
24	7.06			
25	6.49			
26	6.77			
Mean	6.64	7.19	7.47	6.58
Variance	0.0752	0.2106	0.0854	0.1044

Exhibit USPS-T-2D – Mystic Color Labs – Bootstrap Results
(5,500 Iterations)

A. Random Draw from all pieces

	Sample Size					
	100	200	300	1000	5000	12000
Mean	6.63095	6.62219	6.61898	6.61772	6.61499	6.61487
Variance	0.05843	0.02917	0.01972	0.00604	0.00118	0.00049
Standard Deviation	0.24173	0.1708	0.14043	0.07774	0.03436	0.02224
Precision level	7.15%	5.06%	4.16%	2.30%	1.02%	0.66%

B. Two-Stage Sampling – random draw from all sacks, random draw of pieces in selected sacks

number of sacks	10	13	65	20	20
piece sample rate	1/1	1/1	1/5	1/1	1/2
Mean	6.6172	6.6163	6.6161	6.6157	6.6170
Variance	0.0093	0.0071	0.0024	0.0048	0.0057
Standard Deviation	0.0966	0.0842	0.0494	0.0692	0.0753
Precision level	2.86%	2.49%	1.46%	2.05%	2.23%

Account 25 — Sample Size

A. Random draw from all pieces

	100	200	300	1000	5000	12000
Mean	7.14667	7.13859	7.13617	7.13268	7.13227	7.13155
Variance	0.05104	0.02557	0.01691	0.00515	0.00101	0.00043
Standard Deviation	0.22591	0.15990	0.13002	0.07179	0.03183	0.02074
Precision Degree %	6.2	4.4	3.6	2.0	0.9	0.6

B. Random draw from all sacks

Number of sacks	5	10	20	20
Piece sample rate	1/1	1/1	1/1	1/2
Mean	7.13682	7.13323	7.13297	7.13313
Variance	0.02541	0.01215	0.00573	0.00698
Standard Deviation	0.15939	0.11187	0.07572	0.08353
Precision Degree %	4.4	3.1	2.1	2.3

Account 56 — Sample Size

A. Random draw from all pieces

	100	200	300	1000	5000	12000
Mean	7.45749	7.45053	7.44785	7.44681	7.44717	7.44719
Variance	0.04880	0.02413	0.01666	0.00485	0.00097	0.00041
Standard Deviation	0.22091	0.15541	0.12906	0.06966	0.03121	0.02031
Precision Degree %	5.8	4.1	3.4	1.8	0.8	0.5

nobs=479

B. Random draw from all sacks

Number of sacks	5	10	20	20
Piece sample rate	1/1	1/1	1/1	1/2
Mean	7.45706	7.45148	7.44963	7.44890
Variance	0.01748	0.00852	0.00407	0.00476
Standard Deviation	0.13220	0.09231	0.06381	0.06901
Precision Degree %	3.5	2.4	1.7	1.8

Account 63 — Sample Size

A. Random draw from all pieces

	100	200	300	1000	5000	12000
Mean	6.59390	6.58858	6.58531	6.58437	6.58335	6.58313
Variance	0.03435	0.01717	0.01129	0.00332	0.00068	0.00028
Standard Deviation	0.18535	0.13105	0.10623	0.05765	0.02609	0.01662
Precision Degree %	5.5	3.9	3.2	1.7	0.8	0.5

B. Random draw from all sacks

Number of sacks	5	10	20	20
Piece sample rate	1/1	1/1	1/1	1/2
Mean	6.58352	6.58515	6.58505	6.58515
Variance	0.01814	0.00925	0.00547	0.00542
Standard Deviation	0.13468	0.09618	0.06851	0.07363
Precision Degree %	4.0	2.9	2.0	2.2

Exhibit USPS-T-2F – Bootstrap Results on Annual Revenue Adjustment Factor – Nashua
(based on July-September 1996 P.O. samples, sample size = 50)

		If Adjust Revenues Daily	If Adjust By Current Rule (more than +/- 1.5%)	Bias	Estimated Cost to Postal Service of Adjustment Facto Bias (annual)
Mean		1.0197	1.0186	-0.0011	-$5,463.52
Variance		0.0000	0.0000		
Standard Deviation		0.0013	0.0013		
95% Confidence Interval	lower bound	1.0171	1.0161		
	upper bound	1.0226	1.0208		
Precision (annual)	below mean	0.25%	0.24%		
	above mean	0.28%	0.22%		
Precision (monthly)	below mean	0.87%	0.83%		
	above mean	0.97%	0.77%		

*assuming monthly unadjusted revenue of $400,000

Exhibit USPS-T-2G – Costs with New Sample Design

Firm	Activity	Cost Per Piece	Monthly Cost
Mystic	Weighing	$0.0151	
	Accounting		$893
	Sampling		$2,531
	Total		$3,424
Seattle (per account)	Weighing	$0.0104	
	Accounting		$162
	Sampling		$740
	Total		$902
Nashua	Sampling and Manifest		$4,338
	Accounting		$570
	Total		$4,908

Exhibit USPS-T-2H – Costs with Computerized Sampling

Firm	Activity	Cost Per Piece	Monthly Cost
Mystic	Weighing	$0.0151	
	Accounting		$893
	Sampling		$1,548
	Total		$2,441
Seattle (per account)	Weighing	$0.0104	
	Accounting		$162
	Sampling		$646
	Total		$808